Efficient Algorithm for Conformational Search of Macrocyclic Molecules

CHEUK-SAN WANG*

Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Received 13 November 1995; accepted 16 May 1996

ABSTRACT

A new algorithm, *complementarity*, is developed for conformational search of macrocyclic molecules. The algorithm scans a large number of candidate conformations and energy-minimizes only the promising ones. These candidates can be generated by two operators that construct new conformations from known minima. The candidates have similar bonded-interaction energy as the known minima and possibly lower nonbonded interaction energy. This algorithm is 9 to 11 times faster than the existing methods when tested on two large rings, cycloheptadecane and rifamycin SV.

Introduction

The properties and possible interactions of organic molecules are intimately related to their reachable conformations. The probability of a particular conformation being populated at room temperature is determined by its energy, according to the Boltzmann distribution. Therefore, in applications such as drug activity prediction, one is interested in identifying those conformations with energies very close to the global energy minimum. This is the *conformational search problem*.† In particular, given a potential energy function and a molecule, the problem of conformational search is to find all conformations of the molecule whose energies are sufficiently close (e.g., within 3 kcal/mol) to the global minimum energy for that molecule. Unfortunately, the space of possible conformations has $3N - 6$ dimensions where N is the number of atoms in a molecule. There is an enormous number of local energy minima in this space for even a small molecule with tens of atoms.

The goal of our research is to develop better methods for conformational search of *macrocyclic molecules*, which are ring-like molecules with 10 or more bonds in the ring. There are a number of important classes of macrocyclic molecules; for example, the *macrolides* are important antibiotics[1,2] and *macrocyclic musks* are of importance in perfumery. Furthermore, problems such as the conformational search of protein loops during homology

*Author to whom all correspondence should be addressed. E-mail: wang@ai.mit.edu

†An alternative to complete search of conformation space is described in ref. 26, where the search focuses on finding a set of maximally different conformations that ''cover'' the low-energy conformation space.

CCC 0192-8651 / 97 / 020277-13

modeling[3] can be cast as a conformational search of macrocycles.

Cyclic molecules are challenging targets for conformational search because their low-energy conformations are more constrained than those of acyclic molecules. For acyclic molecules, it is generally easy to construct, without recourse to energy minimization or extensive search, candidate conformations that do not have excessive energies; this is the basis of many conformational search strategies.[4] The same cannot be said for cyclic molecules because of the need to satisfy the *ring closure* constraint. This constraint means that atoms in cyclic molecules (especially those with odd-length rings) do not usually fall on any lattice and that bond angles and torsion angles do not necessarily fall at their minimum energy values. Furthermore, atoms in cyclic molecules are generally quite close to each other and, therefore, simply avoiding collisions is challenging.

Existing general conformational search algorithms take several days on typical workstations to find all low-energy conformations of a medium sized macrocyclic molecule. New ring-specific algorithms (see refs. 5, 6, and 7) can perform much better than these general algorithms but are still very time consuming.

Previous Work

Conformational search methods have been studied for many years. Dale[8] has performed early pioneering work on large rings. Currently, most methods do not distinguish between cyclic and acyclic molecules. Many of these methods consist of some technique for generating candidate conformations which are then subjected to energy minimization. These techniques can be separated into two broad categories: deterministic methods and stochastic methods.

Deterministic search methods explore the conformational space systematically. The classic deterministic method is *internal coordinate tree search*.[9] This method generates candidate conformations for minimization by holding the bond lengths and angles fixed and stepping each torsion angle of the rotatable bonds through a series of possible values (e.g., 0°, 60°, 120°, 180°, . . .). In the worst case, every combination of all torsion angle values must be explored. Suppose there are N rotatable bonds in a molecule and each torsion angle takes d values, then up to d^N conformations may be generated and minimized. Therefore, the computational complexity of the algorithm is exponential in the number of rotatable bonds. This makes the approach infeasible for even relatively small molecules. Note that most of the conformations generated by applying this process to a cyclic molecule would violate severely the ring-closure constraint. These candidates are generally discarded before minimization but, nevertheless, this method is quite wasteful for macrocyclic molecules.

Stochastic techniques have the advantage of being able to generate a random sample of low-energy conformations for molecules that are too flexible to explore systematically. For example, the stochastic version of internal coordinate tree search is the *internal coordinate Monte Carlo search*.[10] In each step of the algorithm, several torsion angles are randomly varied. The resulting conformation becomes the new starting point of minimization. When applied to macrocyclic molecules, the random variation of torsional angles often produces conformations that violate the ring-closure constraint. Those structures cannot be used for minimization. The Monte Carlo step has to be repeated until a suitable structure is found.

One of the simplest stochastic techniques is the *Cartesian stochastic search*.[11] It represents a conformation by the Cartesian coordinates of each atom. The method operates by taking a known conformation and applying limited, random translations (''kicks'') to every atom in the molecule. The resulting conformation becomes the new starting geometry for energy minimization.

Molecular dynamics is a stochastic method that simulates the physical behavior of molecules in a thermal bath. The molecule changes its conformation due to thermal motion. The higher the simulated energy, the broader the conformational sampling. At certain time intervals, the algorithm collects the conformations of the molecule being simulated and minimizes them.

The *distance geometry*[12,13] method represents molecules primarily[‡] as a set of lower and upper bounds on the distances between every pair of atoms. These distance bounds embody bonded constraints, such as bond lengths and angles, as well as nonbonded constraints, such as van der

[‡]Distance geometry information (e.g., a matrix listing the distances between all atom pairs) is unable to distinguish between a structure and its mirror image. A single chiral constraint (at least) must be included to eliminate this degeneracy because the stereochemistry of chiral centers is generally known in advance.

Waals radii. The distance geometry representation captures cyclic molecules as well as acyclic molecules. Conformations are typically generated by sampling distances from within these bounds and then generating candidate conformations that best approximate these sampled distances. The resulting conformations are then energy-minimized.

A review and comparison of the performance of these methods in the context of searching for the low-energy conformers of cycloheptadecane, the 17-carbon unbranched cycloalkane, is found in ref. 4. A number of other approaches to conformational search have been explored in the literature.

A genetic algorithm has been applied to conformational search by Judson et al.[14,15] They found that the genetic algorithm is better than a comparison algorithm on an acyclic molecule. Expert-systemlike approaches[16,17] or directed search[18] study the components of molecules and deduce their conformation with a rule set or heuristics. It is not clear whether these approaches extend to macrocyclic molecules.

In addition to these general conformational search methods, several ring-specific algorithms have been developed. These methods typically consist of operators for modifying existing ring conformations, followed by minimization of the resulting structures. They use operations like corner flapping,[5] edge flipping,[6] and torsion flexing.[7] Guarnieri and Wilson[19] have applied simulated annealing with an exact ring closure algorithm to the same problem. The complementarity method described in this article belongs to this class of algorithms. A later section compares their performance to our new approach.

The main problem with existing approaches to conformational search is their inefficiency. For example, with cycloheptadecane all methods required more than 30 cpu days on a MicroVAX II to find most low-energy conformations.[4] Most of the cpu cycles are spent on minimizing the energy of the candidate conformations. For example, starting from a random conformation, each minimization takes about 1 minute on a SPARCstation 2 under BatchMin v3.5.[20] Cartesian stochastic search, internal coordinate tree search, Monte Carlo search, and distance geometry all require about 10,000 minimizations each for this 51-atom molecule. Therefore, these algorithms would take about 7 days on a SPARCstation 2. These methods do not attempt to generate starting geometries close to a local energy minimum. They spend less than 5% of the time finding starting conformations and more than 95% of the time minimizing the energy function on all starting conformations. No selection of the starting conformations was attempted. Only 2% to 3% of the minimized conformations are useful. Most minimizations result in high energy or duplicated conformations.

This suggests that a greater investment in selecting initial conformations likely to be close to local minima might increase the rate of useful minimization and reduce the overall CPU requirement. One can quickly scan many conformations and minimize only the promising ones. Because each minimization requires thousands of energy function evaluations, if one efficiently scans 1000 conformations and minimizes only one or two of them, the use of time is more balanced. The approach described here will minimize only those conformations that have energy close to a minimum. This will reduce the time for minimization and produce more useful conformations. When a starting conformation has energy much above the global minimum, its energy is not always a good predictor of minimized energy. When the starting conformation is only a few kcal/mol above the global minimum, the initial energy becomes an accurate predictor of minimized energy. This observation is one of the cornerstones of the approach described next. Since our algorithm does not know the true global minimum, it is approximated by the current lowest energy.

Complementarity Approach

This section describes a new approach to conformational search of macrocyclic molecules. We call it the *complementarity* approach. Given some known conformational minima of a macrocyclic molecule, one can apply some operators on these minima and generate many new candidate conformations. The quality of these candidates can be found by a single evaluation of the energy function. If a program chooses to minimize only those candidates that have energy close to the current global minimum and are quite different from known minima, it has a much higher chance of obtaining a useful conformation. The time for minimization is also reduced because of the proximity of the candidate's structure to its local minimum. The operators may not produce complete coverage of the conformational space, but we can use other techniques, like the Cartesian stochastic search, to ensure completeness.

The remaining problem is finding a set of operators that can produce many new, low-energy con-

formations given some known minima. For chain or treelike molecules, one can simply change the torsional angle of some rotatable bonds because these molecules do not have the closure constraint. Finding such operators for macrocyclic molecules is much more difficult. Other researchers have used operators like corner flapping or edge flipping,[6] but the candidates they generate have much higher energy than the global minimum. This is because some local geometries (bond angles) are deformed. Before presenting the operators we use, we present two hypotheses about low-energy conformations that motivate the operators:

1. Low-energy conformations have low-energy substructures.
2. Low-energy conformations share similar substructures.

We have found 256 conformations of cycloheptadecane within 3 kcal/mol of the global minimum using the MM2 force field.[§] Figure 1 shows a histogram of the MM2 energies of all eight-adjacent-bond substructures of these conformations. The mean and standard deviation of the energies are 10.257 and 0.921 kcal/mol, respectively. Figure 1 shows that most of the substructures have similar energy. There are very few substructures with high energy. Similarly, we found 42 low-energy conformations of rifamycin SV. Their eight-adjacent-bond substructures have

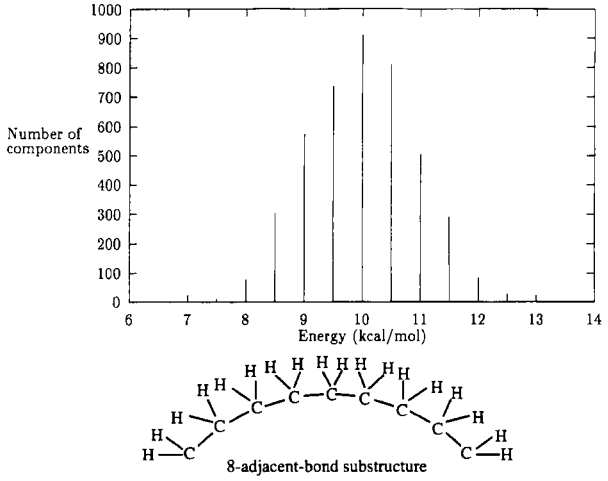

FIGURE 1. Energies of 4352 eight-adjacent-bond substructures of low-energy conformers of cycloheptadecane.

[§]The global minimum of cycloheptadecane's conformational energy is 19.23 kcal/mol.

energies with mean and standard deviation of 15.42 and 1.002 kcal/mol, respectively. These data support the first hypothesis. We postulate that it is generally not the case that a low-energy substructure forms a low-energy conformer with a high-energy substructure with the help of nonbonded interactions.[‖]

The second hypothesis is more subtle than the previous one. We want to show that substructures are ''reused'' in different conformers. We randomly take 300 eight-adjacent-bond substructures from the low-energy conformers of cycloheptadecane. Then we compare the six dihedral angles against the dihedral angles of the other substructures. On average, there are 20.05 other pieces that have angles all within 20° of a selected piece.[¶] This is much higher than would be expected from a uniform distribution of angles. Suppose each dihedral angle can randomly take on one of six values (e.g., 60°, 120°, 180°, . . .). The probability that two pieces would have a similar angle value (within 20°) is $1/6^6 = 0.00002$, whereas the probability for two substructures of different low-energy conformers to have a similar dihedral angle value is about $20.05/(17 \times 256) = 0.0046$. This shows that the distribution of dihedral angles in this molecule is far from random. One can generalize this observation to conformations of different molecules. We postulate that if two different kinds of molecules have a large connected substructure in common, the low-energy conformations of this shared substructure in one molecule will likely appear in the conformations of the other.

Given the above observations, we can use two operators, **combine** and **mirror**, to generate starting conformations with low energy.

COMBINE OPERATOR

The combine operator recombines substructures from conformational minima. We have observed that low-energy conformations have low-energy substructures. New conformations can be found by recombining these substructures. Given a set of conformational minima, we can compute all relative positions of every pair of bonds that are several bonds apart. If two pairs of end bonds in two different substructures have the same relative positions and orientations, then the substructures can

[‖]This statement assumes that the strength of the electrostatic interactions have been appropriately reduced to reflect the presence of a polar solvent.

[¶]The median and standard deviation are 15 and 17.8, respectively. This distribution has a long tail in the positive direction.

substitute for one another (Fig. 2). The bond lengths and bond angles are preserved, but van der Waals interactions can raise or lower the total energy. If the given conformations have a fixed chirality, the new candidates are guaranteed to have the same chirality. Given m conformations of a cyclic molecule with n atoms, $O(m^2n)$ candidates[#] will be examined. If the molecule is symmetric (all substructures have the same chemical formula), there will be $O(m^2n^2)$ candidates to examine; that is, each substructure can replace any other substructure on the ring.

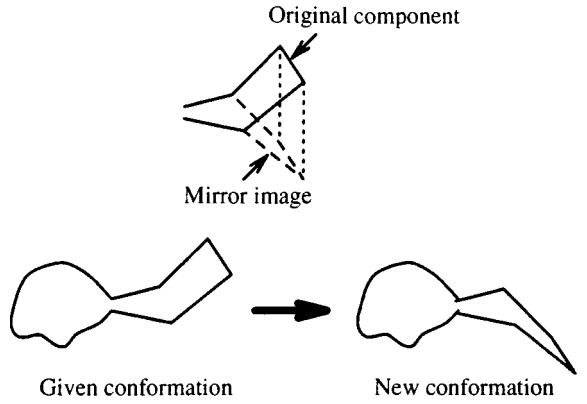

FIGURE 3. The **mirror** operator that replaces a substructure of a conformation by the substructure's mirror image.

MIRROR OPERATOR

The mirror operator substitutes a substructure by its mirror image. Given a conformation, its mirror image (enantiomeric conformation) has exactly the same energy because all distances among atoms are unchanged. We have observed that a low-energy conformation must be composed of low-energy substructures. Hence, the mirror images of the substructures must have low energy. Given a ring conformation, we can partition the molecule into two substructures. If we retain the conformation of one substructure and "glue" on the mirror image of the other substructure, the result would be a very different conformation (Fig. 3). The local geometry of each individual substructure is unchanged. The substructures would have the same energy as before. Additional energy can only arise from bonded interactions at the junctions and new van der Waals interactions between the substructures. We can minimize the change in bonded interactions at the junctions if we retain their local geometries. In other words, we want the bonds at the junctions to have the same bond lengths and angles as the given conformation. This is true if the end bonds of the retained substructure are coplanar and we use the plane of these end bonds as the plane of reflection of the other substructure.

To prove this, consider a conformation consisting of substructures X and Y where A and B are the end bonds of X:

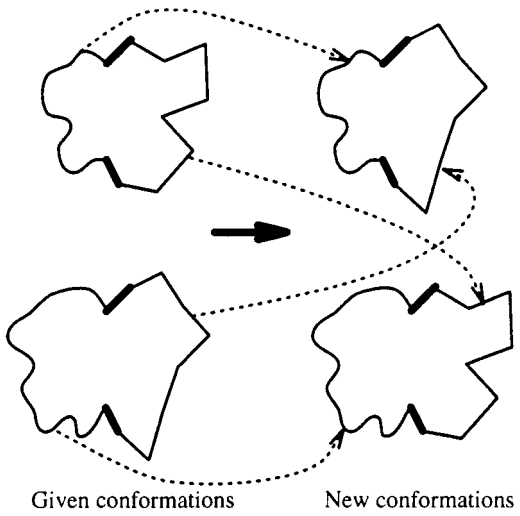

FIGURE 2. The **combine** operator that mixes substructures from two conformations.

Suppose A and B are coplanar. Let R be the operator that reflects a substructure about the AB plane. We use "+" to denote the joining of two substructures and "=" to denote the equivalence between two structures. Clearly, R is distributive over $+$. Because A and B must be on the AB plane, $R(A) = A$ and $R(B) = B$. Therefore, $R(Y + A + B) = R(Y) + R(A) + R(B) = R(Y) + A + B$. This says that reflecting Y produces the structure as reflecting Y and A and B. Therefore, the bond lengths and angles at the junction are unchanged

[#] $O(f(n))$ represents a function with order of magnitude $f(n)$. A function, $g(n)$, is said to be $O(f(n))$ if $g(n) \leq cf(n)$ for some constant, c.[27]

after replacing Y by the mirror image of Y. The torsional angles at A and B may be different but they introduce very little energy. There would be new van der Waals interactions which can increase or decrease the total energy.

Notice that a substructure would have the opposite chirality of its mirror image. To preserve the chirality of new conformations, the reflected substructure cannot have any chiral center.

The mirror operator is a specialized form of the combine operator. It is equivalent to applying combine to a conformation and its mirror image. Goto's operators are special cases of this operator. Corner flapping[5] reflects two bonds, whereas edge flipping[6] reflects three bonds. By looking at all pairs of bonds and computing their planarity, we explore more conformational space more efficiently. There is also a greater chance of lowering total energy because more bonds are changed. A later section compares the complementarity approach with Goto's algorithm.

Can the operators be applied to substructures of any size? With a few approximations, we can find the limitation to the sizes of substructures. We assume that all bond lengths and bond angles are fixed, but the dihedral angles are free to change. Go and Scheraga[21] have shown that two fixed end bonds could generate six nonlinear equations on the dihedral angles. Therefore, at least six dihedral angles are needed to satisfy the equations. To apply the operators, the smallest substructure would have six bonds (Fig. 4). As a consequence, the smallest ring on which we can apply the operators has two minimal size substructures and $6 + 6 - 2 = 10$ bonds.

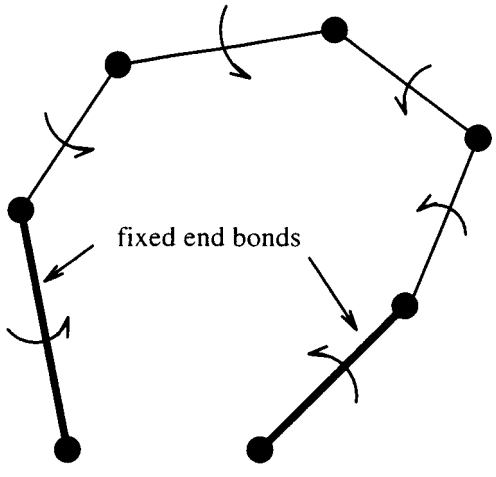

FIGURE 4. A minimal size substructure.

The Algorithm in Detail

The algorithm makes use of two data structures: a priority queue and a database of substructures. The priority queue** stores the starting conformations prioritized by their energy values. The lowest energy starting conformation can be removed and new ones can be added very efficiently. This data structure allows us to find and minimize the best starting conformation first.

The other data structure, a database, stores substructures of conformational minima. It is indexed by the type of substructures (their atoms and bonds), and the relative orientation of end bonds. In the current implementation, the relative orientation is encoded with the distances among atoms forming the end bonds†† (Fig. 5). The database is implemented as a hash table.‡‡ The hash function applies to the types of substructures and a discretization of the distances. Given a substructure of a conformer, we can find all of its *complementary*

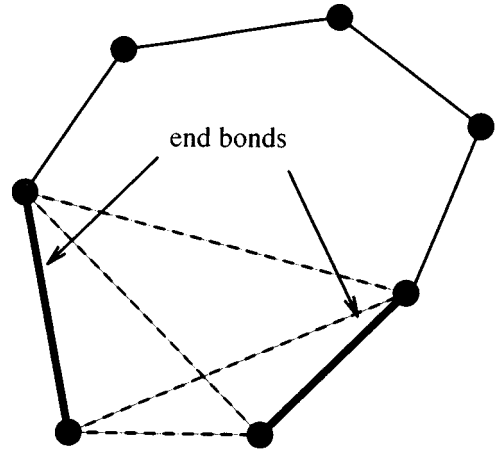

FIGURE 5. Four distances (indicated by dotted lines) used to encode the relative configuration of the end bonds.

**A priority queue is a set that supports the *Insert*, *FindMin*, and *DeleteMin* operations. Each element in the set is associated with a key (a number). *Insert* adds an element and its key to the queue. *FindMin* and *DeleteMin* finds or removes the element with the smallest key in the set. A priority queue can be implemented efficiently as a heap.[27]

††This encoding does not distinguish between a substructure and its enantiomer. We make use of this fact so as not to store the enantiomers.

‡‡A hash table is like a dictionary. The table supports the *Insert*, *Member*, and *Delete* operations. Each element in the table is associated with a key. The *hash function* computes the location of an element from its key. The *Member* operation finds all elements in the table with a certain key. All operations on the hash table can be performed in a constant amount of time.[27]

substructures from the database efficiently. A substructure's complement is another substructure with the complementary bonds in the ring, and whose end bonds have the same relative orientation (Fig. 6). To find the complements of a substructure, we simply look up the complementary bonds and apply the hash function.

With these data structures, we define a procedure **Generate_starting_conformations** that finds starting conformations from a newly discovered minimum. Basically, it systematically applies the operators described previously to the new conformation.

Generate_starting_conformations (conformation) proceeds as follows. For all connected substructures C with six or more single bonds in the conformation, we do the following:

1. If C is already in the substructure database, stop.
2. Otherwise, add C to the database.
3. Find all complementary substructures of C from the database.
4. Find transformations of the complementary substructures such that C and the substructures would form new starting conformations and the ring closure constraint is satisfied.
5. Compute the energies of the starting conformations.

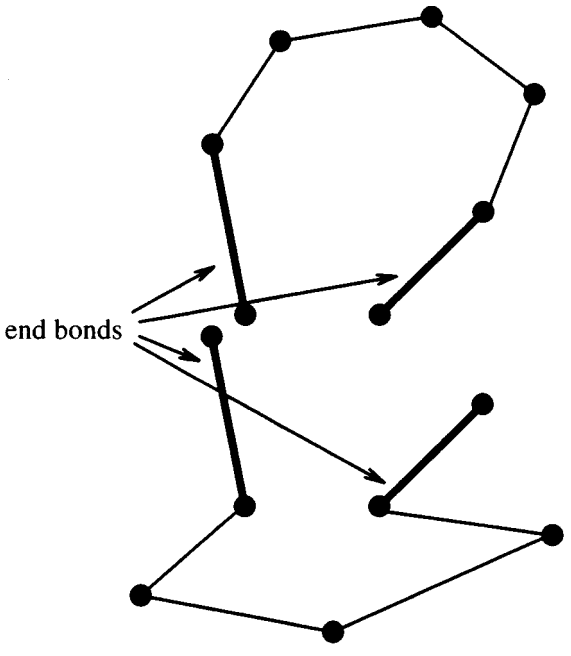

FIGURE 6. Two substructures that are complements of each other.

6. Add the starting conformations that have energy within ΔE of the global minimum to the priority queue. ΔE must be bigger than the ultimate energy window and depends on the complexity of the molecule. If the molecule is small and symmetric, 5 kcal/mol is sufficient. If the molecule is large and asymmetric, ΔE can be as high as 15 kcal/mol. One can overestimate ΔE because this only enlarges the priority queue.

In the current implementation, the conformations are represented by the Cartesian coordinates of the atoms. The end bonds of C and its complementary substructures have the same relative configurations, but their Cartesian coordinates do not necessarily match. To form a new starting conformation, a rigid transformation (step 4) is needed to transform the end bonds and other atoms of the complementary substructure. This is achieved by the algorithm in ref. 22. A 4×4 transformation matrix can be computed to best match (in the least-squares distance sense) the coordinates of the atoms on the end bonds of C and the complementary substructure.

It may seem that only the combine operator, which recombines substructures from conformational minima, is used in this procedure, but there is another twist to step 4. A substructure can be transformed to its mirror image by a transformation matrix whose determinant is -1. This matrix is also computed and returned if it can produce a close match of the end bonds. Thus, the mirror image of a complementary substructure is used to generate a starting conformation. Because the two halves of a conformer are always complements of each other, the mirror operator is also applied at this step.

With the procedure that generates starting conformations, we can describe the top level algorithm:

1. A few conformational minima (typically below 100) are found by a randomized method such as Cartesian stochastic search or internal coordinate Monte Carlo search. These minima should have the correct chirality and be dispersed in the conformational space. If an internal coordinate Monte Carlo search is used, six or more torsion angles should be changed at once.
2. The procedure **Generate_starting_conformations** is called with these minima as arguments. The main purpose of this step is to fill

the substructure database with some entries. This is the ''seeding'' phase.

3. The starting conformation with the lowest energy before minimization is removed from the priority queue. If the queue is empty, a Monte Carlo operation is performed instead.
4. If the starting conformation is sufficiently different from the known minima, it is minimized. This check is necessary because many similar starting conformations are found by the algorithm.
5. If the new minimum has not been found before, **Generate_starting_conformations** is called with it as the argument.
6. Go to 3.

This algorithm has been implemented in Common Lisp.[23] The MM2 force field is used for minimization and energy evaluation. The next section will evaluate the performance of the algorithm.

Performance Evaluation

We compared the new algorithm with several existing methods, using the number of energy minimizations and the number of ''useful'' conformational minima found. We chose the number of minimizations instead of cpu time because the former is independent of computer hardware or programming languages. For existing algorithms, most of the resources are used for minimization.

CYCLOALKANE

The cycloalkanes have a regular ring structure of $(CH_2)_n$. Their conformations have been studied extensively. They are good benchmarks for evaluating conformational search algorithms. Since cycloalkanes have no polar atoms, the nonbonded interactions are limited to van der Waals forces. Additionally, their structures are completely symmetric; all substructures with the same number of bonds are of the same type in the database. More starting conformations could be generated because our algorithm exploits this symmetry. No other method can make explicit use of this symmetry. Because of these features of cycloalkanes, the complementarity algorithm vastly outperforms existing methods.

For the following results on cycloalkane, only 30 minima are found using the randomized method in step 1 of the top level procedure.

Our program is first run on cyclotridecane, the 13-carbon cycloalkane, to validate its completeness. We use BatchMin v3.5[20] and its MM2 force field for minimization. ΔE is set to 6 kcal/mol. The global minimum is found to be 20.415 kcal/mol. Figure 7 shows the results of the new algorithm and Cartesian stochastic search[11] running on the molecule. After 2000 Monte Carlo steps, 15 conformers are found to be within 3 kcal/mol of the global minimum. We believe that these are all the conformers because every minimum has been found several times. A Cartesian stochastic search finds all useful conformations with 341 minimizations against 98 minimizations for the complementarity algorithm. The complementarity method is 3.48 times faster. More importantly, the result suggests that the algorithm completely explores the low-energy conformational space. Table I compares the performance of the complementarity algorithm to Goto and Osawa's corner-flapping algorithm.[5] Our algorithm has a higher ratio of low-energy conformations to total number of conformations found.

The second cycloalkane used as a benchmark is cycloheptadecane, the 17-carbon cycloalkane. Saunders et al.[4] applied seven methods to search for its conformations. They found that several of the methods have roughly similar performance. We applied the complementarity algorithm to this molecule using BatchMin v3.5 and the MM2 force field. These are the exact program and energy functions used in Saunders et al.'s study.[4] We can directly compare our data with those in their work. Our program found the same global minimum (19.23 kcal/mol) as several other methods in their article, but the complementarity algorithm did this after only 41 energy minimizations. Figure 8 shows the performance of the algorithm using MM2. All minima are tested by BatchMin's normal mode analysis. Table II is a comparison of the best method in their study, usage-directed torsional Monte Carlo search, against our algorithm. Comparing the number of minimizations each algorithm needs to find the 203rd conformation, in the 3-kcal/mol bracket, where about 80% of the useful conformers are found, the complementarity algorithm is 9.4 times faster than usage-directed torsional Monte Carlo search, likewise, it is 12.2 and 14.7 times faster in finding the 232nd and 249th useful conformation, respectively.[§§]

[§§] Ref. 4 describes the cpu time for finding the 203rd, 232nd, and 249th useful conformation. We deduced the approximate number of minimizations from their data.

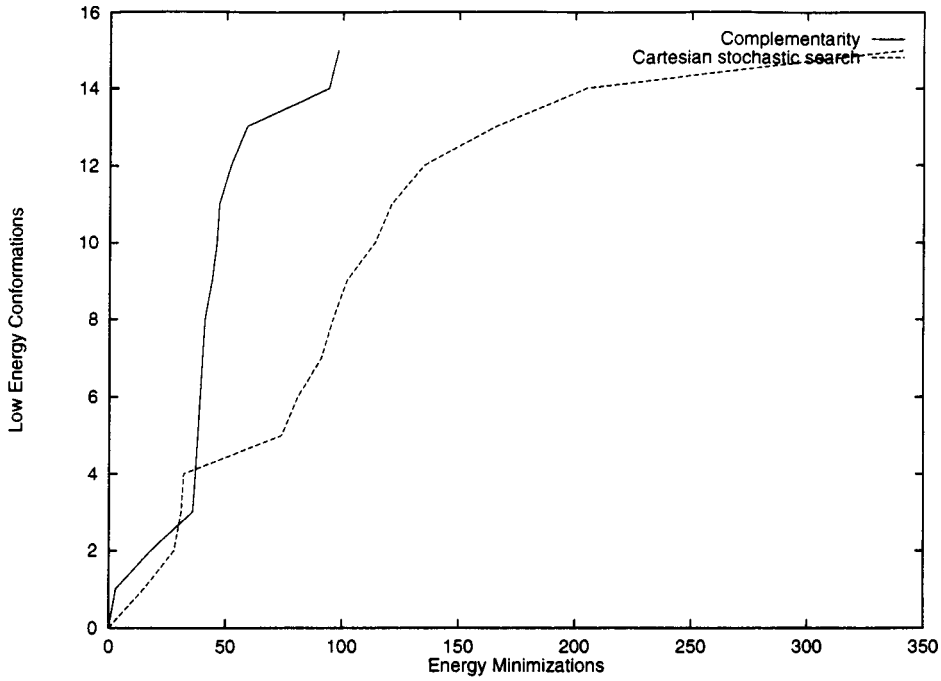

FIGURE 7. Performance of Cartesian stochastic search and the complementarity algorithm on cyclotridecane.

Figure 8 also shows that conformers within 1 and 2 kcal/mol of the global minimum are found very rapidly in the beginning of the search. In addition, at the early stage of the run, nearly every minimization produces a useful conformation. Table III is a comparison of the rate of conformational search during the early stage. The initial rate of conformation discovery is much higher using the complementarity algorithm. This is because the operators are very effective in finding candidates that minimize to very low energy and because the priority queue allows us to minimize the starting conformations in order of their energy. Thus the energies of the minima are roughly in increasing order. Figure 9 illustrates this point. The sharp peaks in the graph are caused by the Monte Carlo steps of the algorithm (step 1 and sometimes step 3 of the top level procedure). The rest of the minima have a slow upward trend in energy. This trend shows that the best minima are likely to be found early in the search. In fact, the conformation of cycloheptadecane with the lowest energy is found after only 41 energy minimizations.

RIFAMYCIN SV

Our algorithm performs very well on cycloalkanes. To evaluate our method more completely, we try it on a radically different molecule, rifamycin SV. It is a well-known representative of the ansamycin family[24] (Fig. 10). Kolossvary and Guida[7] have also worked on the conformations of this molecule. They suggest that it is an extremely difficult conformational search problem for current

TABLE I.
Performance of Corner Flapping[a] and the Complementarity Algorithm on Cyclotridecane.

Method	Energy minimizations	Number of conformations generated	Number of conformers having a Boltzmann distribution larger than 0.01%
Corner flapping	Not reported	116	44
Complementarity	98	59	38

[a] Data from corner flapping from ref. 6.

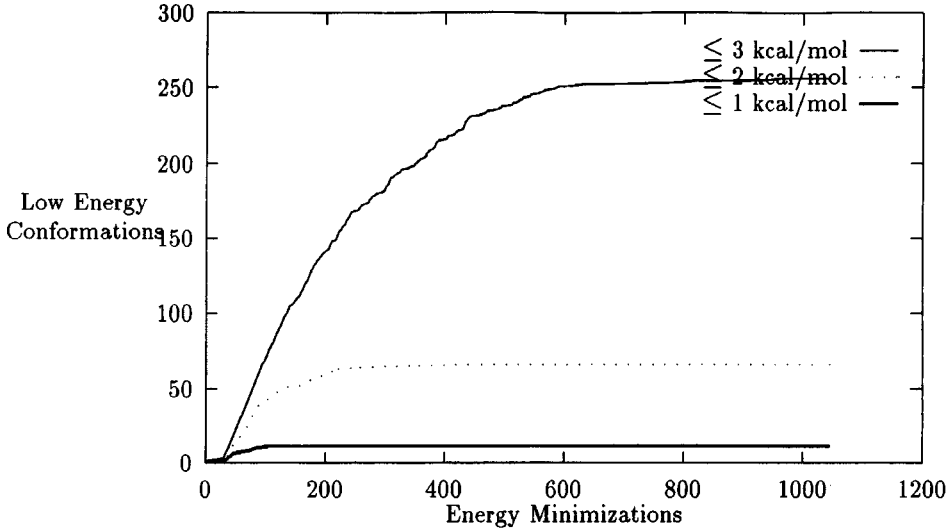

FIGURE 8. Performance of the complementary algorithm on cycloheptadecane with the MM2 force field.

technology. The molecule is completely asymmetric. There are strong electrostatic interactions and several intramolecular hydrogen bonds. Our algorithm cannot exploit any symmetry as in cycloalkane, but the method still outperforms others.

We use the MM2 force field and BatchMin v3.5 for the search. The dielectric constant was set to 10 so as to attenuate electrostatic interactions. These are the same parameters used by Kolossvary and Guida. One hundred minima are found using the randomized method in step 1 of the top level procedure. ΔE is set to 15 kcal/mol. Only the combine operator was used because all substructures have at least one chiral center. The lowest energy conformer found by our algorithm has an energy of 50.006 kcal/mol, which is lower than the 55.69 kcal/mol found by Kolossvary and Guida. Figure 11 shows the result of running our algorithm for 892 energy minimizations. Table IV is a comparison of methods for conformational search of rifamycin SV. Complementarity found the 42nd conformer within 3 kcal/mol of the global mini-

TABLE II.
Unique Conformers Found, Within δE kcal / mol Above Global Minimum, Versus Energy Minimizations During Conformational Searches of Cycloheptadecane.

δE			Number of energy minimizations			
kcal / mol	359	847	1045	3388	5647	8471
			Number of unique conformations			
			Still / Chang / Guida usage-directed torsional Monte Carlo search[a]			
1		10		10	11	11
2		44		61	66	69
3		110		203	232	249
			Complementarity			
1	11	11	11			
2	65	66	66			
3	203	254	256			

[a]Data on usage-directed torsional Monte Carlo search from ref. 5.

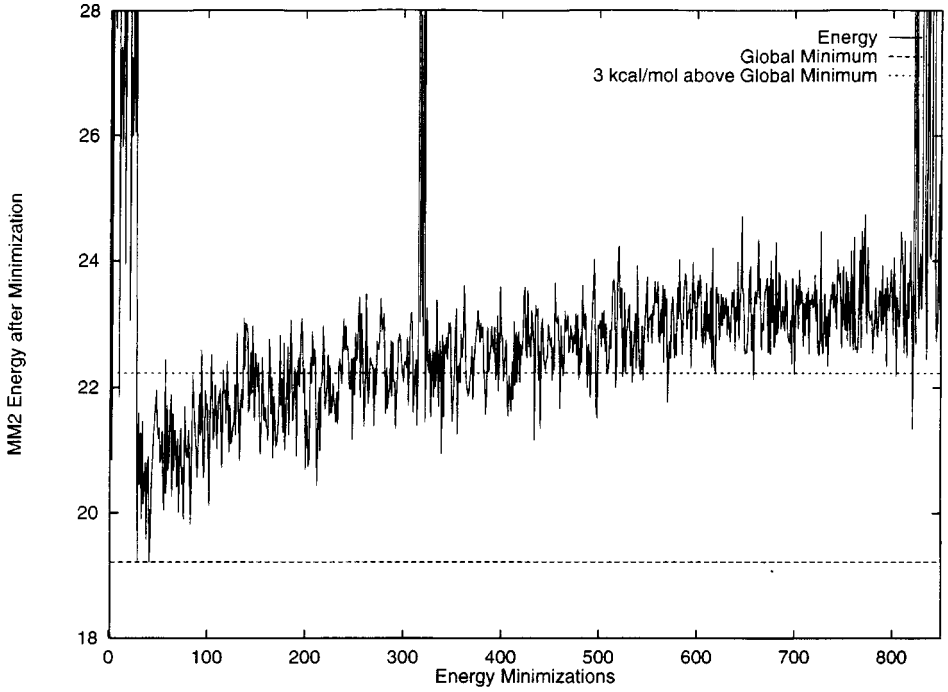

FIGURE 9. MM2 energies of cycloheptadecane conformers after minimizations. The sharp peaks in the graph are caused by the Monte Carlo steps of the complementarity algorithm.

TABLE III.
Rate of Conformational Search of Cycloheptadecane.

Method	Percentage of total minima found / 100 starting geometries		
	1 kcal / mol	2 kcal / mol	3 kcal / mol
Usage-directed torsional Monte Carlo Search	9.1	6.4	4.4
Complementarity	90.9	59.4	26.3

mum after 849 minimizations versus 10,000 for FLEX, the best of the other algorithms. It is 11.8 times faster than FLEX.

Conclusion

This study has described a new algorithm, complementarity, for the conformational search of macrocyclic molecules. The algorithm scans a large number of candidate conformations and minimizes only the promising ones. These candidates can be generated by two operators that construct new conformations from known minima. The can-

TABLE IV.
Unique Conformers Found, Within δE kcal / mol Above Global Minimum, Versus Energy Minimizations During Conformational Searches of Rifamycin SV.[a]

δE	Number of Energy Minimizations			
kcal / mol	892	1000	2000	10000
	Number of unique conformations			
		SUMM		
3				39
6		63		186
		FLEX		
3				42
6		70		305
		Usage-directed torsional Monte Carlo search		
3			0	
6			17	
		Complementarity		
3	42			
6	227			

[a]Data on SUMM and FLEX are from ref. 8.

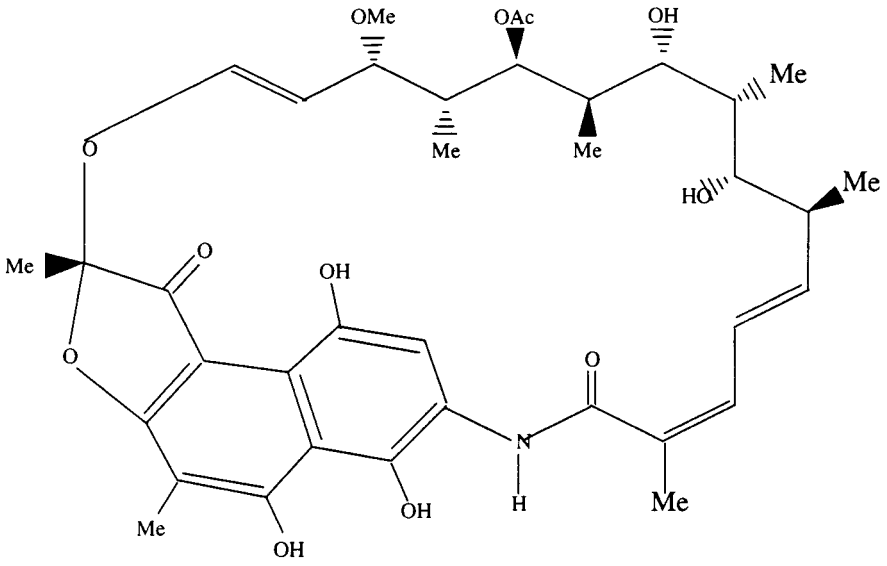

FIGURE 10. Chemical formula of rifamycin SV.

didates have similar bonded-interaction energy as the known minima and possibly lower nonbonded interaction energy. The substructures of conformers are accessed efficiently from a database. The starting conformations are ordered by energy in a priority queue. With cycloheptadecane and rifamycin SV, this algorithm is 9 to 11 times faster than the existing methods.

There are several reasons for the efficiency of the algorithm:

1. The most important reason is that the operators are able to generate good conformations close to the local minimum. Thus, we can use initial energy of the conformations for selec-

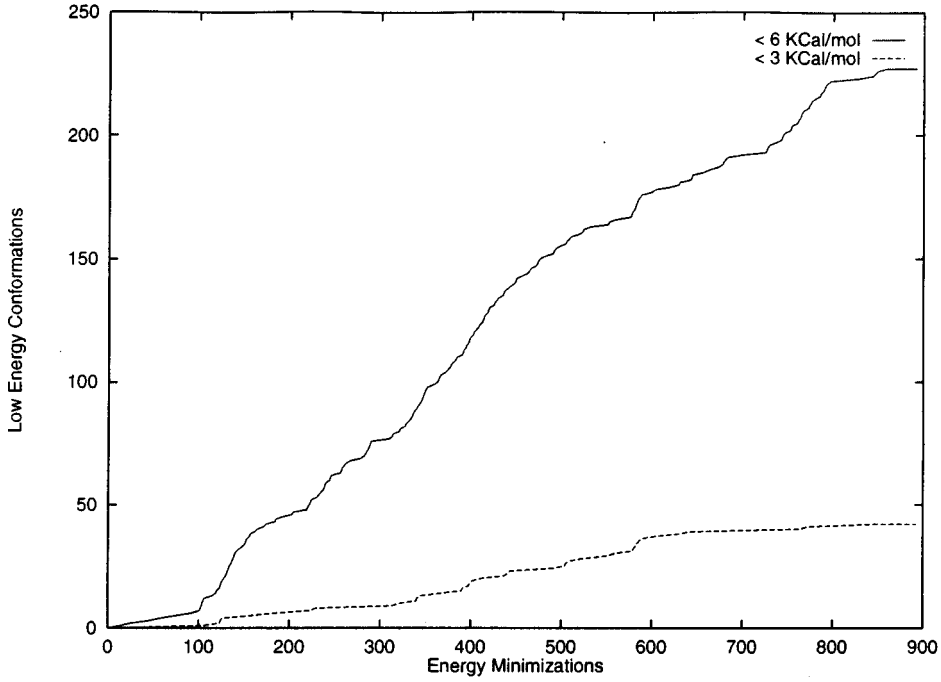

FIGURE 11. Performance of the complementarity algorithm on rifamycin SV.

tion. Low-energy starting conformations increase the chance of finding useful minima and reduce the time of each minimization.

2. The database allows for the systematic retrieval of the complements of any substructure. The operators can be applied systematically to generate good starting conformations. There is no redundancy in generation.
3. The priority queue allows us to minimize the starting conformations in order of their energy. Thus, lowest energy minima are likely to be found early in the search.

On the other hand, the complementarity algorithm may not work well with molecules with multiple long range interactions. The operators do not maintain favorable long range interactions like hydrogen bonds between substructures. Consequently, candidates without these favorable interactions will be generated and rejected because of higher energy.

For rings with long ''side-chains,'' the complementarity algorithm may not explore the side-chains' conformational space fully. One may use our algorithm to find conformations of the ring and a different algorithm for conformations of the side-chains.

The complementarity algorithm may be modified to find conformations of loops in proteins. The ends of a loop are assumed to be fixed (anchored). An operator similar to **combine** recombines peptides from different conformational minima. The starting conformational minima are selectively minimized. This approach may be superior to other methods on conformational search of extended loops.

Furthermore, we have shown that initial energy is a good predictor of minimized energy if the starting conformations of macrocyclic molecules satisfy the bond length and angle constraints. This may also apply to acyclic molecules. We believe that if the starting conformations of acyclic molecules are generated carefully and selectively minimized based on their initial energy, their low-energy conformations can also be found efficiently.

Acknowledgments

The author thanks Professor Tomás Lozano-Pérez for support and discussions: Professor Bruce Tidor, Carlo Maley, and Kevin Lew for comments on the manuscript; and Tao Ke and Jeffrey Song for discussions on stereochemistry.

References

1. S. Omura, *Macrolide Antibiotics, Chemistry, Biology, and Practice*, Academic Press, Orlando, FL, 1984.
2. A. J. Bryskier, J.-P. Butzler, H. C. Neu, and P. M. Tulkens, Eds., *Macrolides, Chemistry, Pharmacology and Clinical Uses*, Arnette Blackwell, Paris, France, 1993.
3. R. E. Bruccoleri and M. Karplus, *Biopolymers*, **26,** 137 (1987).
4. M. Saunders, K. N. Houk, Y.-D. Wu, W. C. Still, M. Lipton, G. Chang, and W. C. Guida, *J. Am. Chem. Soc.*, **112,** 1419 (1990).
5. H. Goto and E. Osawa, *J. Am. Chem. Soc.*, **111,** 8950 (1989).
6. H. Goto and E. Osawa, *J. Chem. Soc. Perkin Trans.*, **II,** 187 (1993).
7. I. Kolossvary and W. C. Guida, *J. Comput. Chem.*, **14,** 691 (1993).
8. J. Dale, *Acta Chem. Scand.*, **27,** 1115 (1973).
9. M. Lipton and W. C. Still, *J. Comput. Chem.*, **9,** 343 (1988).
10. G. Chang, W. E. Guida, and W. C. Still, *J. Am. Chem. Soc.*, **111,** 4379 (1989).
11. M. Saunders, *J. Am. Chem. Soc.*, **109,** 3150 (1987).
12. G. M. Crippen and T. F. Havel, *Distance Geometry and Molecular Conformation*, Chemometrics Series, Wiley, New York, 1988.
13. P. K. Weiner, S. Profeta, Jr., G. Wipff, T. Havel, I. D. Kuntz, R. Langridge, and P. A. Kollman, *Tetrahedron*, **39,** 1113 (1983).
14. D. B. McGarrah and R. S. Judson, *J. Comput. Chem.*, **14,** 1385 (1993).
15. R. Judson, E. Jaeger, A. Treasurywala, and M. Peterson, *J. Comput. Chem.*, **14,** 1407 (1993).
16. D. P. Dolata, A. R. Leach, and K. Prout, *J. Comput.-Aid. Molec. Des.*, **1,** 73 (1987).
17. Z. Ai and Y. Wei, *J. Chem. Informat. Comput. Sci.*, **33,** 635 (1993).
18. A. R. Leach and K. Prout, *J. Comput. Chem.*, **11,** 1193 (1990).
19. F. Guarnieri and S. R. Wilson, *Tetrahedron*, **48,** 4271 (1992).
20. *BatchMin Documentation*, Department of Chemistry, Columbia University, New York, NY 10027, April 1990.
21. N. Go and H. A. Scheraga, *Macromolecules*, **3,** 178 (1970).
22. D. R. Ferro and J. Hermans, *Acta Crystallog.*, **A33,** 345 (1977).
23. G. L. Steele, Jr., *Common Lisp*, 2nd ed., Digital Press, Bedford, MA, 1990.
24. S. K. Arora, *Molec. Pharmacol.*, **23,** 133 (1983).
25. R. Sayle, *RasMol v2.4, A Molecular Visualisation Program*, Biomolecular Structure, Glaxo Research and Development, Greenford, Middlesex, UK, June 1994.
26. A. Smellie, S. L. Teig, and P. Towbin, *J. Comput. Chem.*, **16,** 171 (1995).
27. A. Aho, J. E. Hopcroft, and J. D. Ullman, *Data Structures and Algorithms*, Addison-Wesley, Reading, MA, 1982.